Published online 16 May 2018

Nucleic Acids Research, 2018, Vol. 46, No. 14 ***7339–7353***
doi: 10.1093/nar/gky361

The landscape of the non-canonical RNA-binding site of Gemin5 unveils a feedback loop counteracting the negative effect on translation

Rosario Francisco-Velilla[1], Javier Fernandez-Chamorro[1], Ivan Dotu[2,3] and Encarnación Martinez-Salas[1,*]

[1]Centro de Biología Molecular Severo Ochoa, CSIC-UAM, Nicolás Cabrera 1, 28049 Madrid, Spain, [2]Pompeu Fabra University (UPF), 08003 Barcelona, Spain and [3]IMIM - Hospital del Mar Medical Research Institute, 08003 Barcelona, Spain

Received November 10, 2017; Revised April 13, 2018; Editorial Decision April 19, 2018; Accepted May 08, 2018

ABSTRACT

Gemin5 is a predominantly cytoplasmic protein that downregulates translation, beyond controlling snRNPs assembly. The C-terminal region harbors a non-canonical RNA-binding site consisting of two domains, RBS1 and RBS2, which differ in RNA-binding capacity and the ability to modulate translation. Here, we show that these domains recognize distinct RNA targets in living cells. Interestingly, the most abundant and exclusive RNA target of the RBS1 domain was Gemin5 mRNA. Biochemical and functional characterization of this target demonstrated that RBS1 polypeptide physically interacts with a predicted thermodynamically stable stem–loop upregulating mRNA translation, thereby counteracting the negative effect of Gemin5 protein on global protein synthesis. In support of this result, destabilization of the stem–loop impairs the stimulatory effect on translation. Moreover, RBS1 stimulates translation of the endogenous Gemin5 mRNA. Hence, although the RBS1 domain downregulates global translation, it positively enhances translation of RNA targets carrying thermodynamically stable secondary structure motifs. This mechanism allows fine-tuning the availability of Gemin5 to play its multiple roles in gene expression control.

INTRODUCTION

Post-transcriptional mechanisms governing gene expression depend on the concerted action of RNA-binding proteins (RBPs) and RNAs. Soon after transcription commences, the mRNA associates to distinct RBPs giving rise to dynamic ribonucleopotein (RNP) entities, which control gene expression as a function of the factors present in the complex (1). RBPs typically comprise one or more known RNA-binding domains (RBD), in addition to protein–protein interaction modules (2). However, a significant number of recently discovered RBPs do not contain conventional RBDs (3).

Gemin5 is a peripheral protein of the survival of motor neuron (SMN) complex in metazoan organisms (4–6). This multi-protein complex plays a critical role in the biogenesis of small nuclear ribonucleoproteins (snRNPs), the components of the splicing machinery (7,8). Gemin5 is responsible for recognition of the Sm site of snRNAs, and delivers these molecules to the SMN complex in the cytoplasm (9,10). The protein Gemin5 contains distinct functional domains. At the N-terminus, a WD40 repeat domain is responsible for the delivery of the SMN complex to snRNAs (11–13). In contrast, the C-terminal domain harbors a non-canonical bipartite RNA-binding site consisting of RBS1 and RBS2 domains (14,15). RBS1 and RBS2 domains differ in RNA-binding capacity, and also in the ability to modulate IRES-dependent translation. Moreover, NMR structural analysis of the RBS1 polypeptide showed a mixture of conformations in solution, commonly found in unstructured domains of proteins. Conversely, RBS2 was predicted to contain several helices, of which one was leucine-rich and another glutamine-rich. Thus, separate protein regions are involved in the recognition of RNAs with different functions, primary sequence, and structural organization. These distinctive features strongly suggest the existence of multiple RNA targets recognized by each of these specialized domains likely involved in different functional complexes.

Beyond its role in snRNPs assembly, Gemin5 was identified as a negative regulator of translation (16). However, other laboratory reported a stimulatory effect of the full-length protein for the SMN mRNA, which is predicted to adopt a complex secondary structure (17). The region of the protein involved in this effect remains elusive. We have shown that the endogenous Gemin5 protein sediments with

*To whom correspondence should be addressed. Tel: +34 911964619; Fax: +34 911964420; Email: emartinez@cbm.csic.es

the polysome fractions (18), and that the purified protein interacts directly with ribosomal particles via its N-terminal domain. These data strongly suggest that Gemin5 may control global protein synthesis through its direct binding to the ribosome. In addition, an unbiased proteomic approach identified a large number of cellular factors differentially associated to N-terminal domain of Gemin5, mostly bound through RNA bridges (18). However, the cellular RNA targets of RBS1 and RBS2 domains of Gemin5 remain unknown. Neither the function of these domains in the expression of cellular mRNAs is known.

Understanding the complexity of Gemin5 function in gene expression pathways would greatly benefit from a global approach to identify its targets in the cellular context. Here we have undertaken the challenge to identify the RNAs associated to RBS1 and RBS2 domains in living cells using a genomic approach. The results suggested that Gemin5 acts as a platform, serving as a hub for distinct RNA-protein networks. Interestingly, among the RNA targets of RBS1, the most abundant hit was Gemin5 mRNA. Biochemical and functional characterization of this RNA target demonstrated that RBS1 physically interacts with its own mRNA. This interaction provides a regulatory feedback loop that results in counteracting the negative effect of Gemin5 on translation control, as illustrated here for a specific mRNA region.

MATERIALS AND METHODS

Constructs

The sequences encoding the RBS1 and RBS2 domains of Gemin5 (18) were inserted into pcDNA3-CTAP (19) to generate the constructs pcDNA3-RBS1-CTAP and pcDNA3-RBS2-CTAP using standard procedures. The constructs pGEM3-H12, pGEM3-H34, pGEM3-H1 and pGEM3-H2 were obtained inserting the sequences H12, H34, H1 and H2 from pcDNA3-Xpress-G5 into pGEM3. The plasmid pCAP-luc was generated in two steps. First, the BamHI site was substituted by EcoRI site in plasmid Tagged-IRES (Addgene plasmid # 35570) (20) using the QuikChange mutagenesis procedure (Agilent Technologies) with the primers 5′CAP-luc and 3′CAP-luc. Then, ligation of the EcoRI digested plasmid generated the pCAP-luc construct. The plasmid pCAP-H12-MS2h was generated removing the EcoRI-BamHI fragment of Tagged-IRES construct and inserting the H12 sequence. The plasmid pCAP-luc-EcoRV was generated by mutagenesis to introduce an EcoRV site in pCAP-luc using the QuikChange mutagenesis procedure (Agilent Technologies) with the pair of primers 5′EcoRV-CAP-mut and 3′EcoRV-CAP-mut. The constructs pCAP-luc-H12 and pCAP-luc-H34 were generated inserting the sequences H12 and H34 from pcDNA3-Xpress-G5 into pCAP-luc-EcoRV. The constructs pGEM3-H12d and pCAP-H12d-MS2h were prepared ligating the pair of primers 5′H12-mut and 3′H12-mut, digesting with the indicated restriction enzymes and inserting into the plasmid pGEM3 and pCAP-luc, respectively. Finally, the construct pCAP-luc-H12d was generated inserting the sequence H12d into pCAP-luc-EcoRV. Oligonucleotides (Sigma) used for PCR and restriction enzyme sites used for cloning are described in Supplementary Table S1. All plasmids were confirmed by DNA sequencing (Macrogen).

Cross-link RNA affinity-tandem (CRAFT) purification

HEK293 cells (4×10^7) were transfected with the plasmids pcDNA3-CTAP, pcDNA3-RBS1-CTAP and pcDNA3-RBS2-CTAP using lipofectamine (Thermo Scientific) according to the manufacturer's instructions. Monolayers were washed 24 h post-transfection with ice-cold PBS prior to UV irradiation (150 mJ/cm^2 at 254 nm). Since the RNA-binding properties of Gemin5 indicated that RNA recognition depends upon RNA structure (10,15,17), RNase treatment step was omitted from the UV-crosslinked complexes purification. This procedure will be designated CRAFT for crosslink-RNA affinity-tandem. The rationale was to obtain information on the recognition of the target, which presumably contains a folded RNA region longer than the protein-binding motif usually determined in conventional CLIP procedures (21,22). The RNA-protein complexes associated to RBS1, RBS2 or the TAP polypeptide were isolated as in the tandem affinity purification procedure until the second purification step (23). After washing the calmodulin resin with the associated complexes, beads were incubated with 500 μl of Proteinase K Buffer (10 mM Tris–HCl pH 8, 100 mM NaCl, 0.5% SDS, 2 mM EGTA, 1 mM EDTA, 0.2 mg/ml proteinase K) at 65°C, during 2 h with gentle rocking. The elution products were recovered by gravity flow. RNAs were then extracted with phenol-chloroform, ethanol precipitated and resuspended in 20 μl of H_2O. Each RNA preparation was monitored using the Agilent 2100 Bioanalyzer (Agilent technologies).

cDNA libraries preparation and Illumina sequencing

RNAs (∼100 ng) obtained from two independent biological replicates of each sample (RBS1, RBS2 and TAP) were used to generate cDNA libraries using the TruSeq Stranded mRNA Sample Prep Kit v2 (Illumina) according to the manufacturer's protocol, but omitting the polyA+ selection step. All purification steps were done using AMPure XP beads (Beckman Coulter). cDNA libraries were subjected to Illumina adapter ligation with TruSeq Universal Adapter 5′AATGATACGGCGACCACCGAGATCTACACTCT TTCCCTACACGACGCTCTTCCGATCT3′ and TruSeq Adapter, Index 5′GATCGGAAGAGCACACGTCTGA ACTCCAGTCACATCTCGTATGCCGTCTTCTGC TTG3′ and were sequenced in an Illumina HiSeq2000 sequencer. Reads were 50 bp in a single-read run cycle. The total number of reads obtained for each sample is shown in Supplementary Figure S1A.

Computational analysis

Peaks (compiled reads) for RBS1, RBS2 and TAP replicates were calculated using pyicoclip software (24). For the TAP samples, the peaks of both replicates 1 and 2 were combined and used as a single TAP control to discard any possible false positive. Then, peaks in each replicate of RBS1 and RBS2 that appear in the TAP peaks were removed (Supplementary Figure S1A). The overlapping peaks of the RBS1

or RBS2 replicates were considered as the final set of peaks using fjoin.py (25). Following addition of 20 flanking nts, the peaks were annotated, and peaks assumed to be in intergenic regions were removed (GRCh37/hg19) (Dataset 1).

To verify if the RBS1 and RBS2 peaks had reads supported by the full-length Gemin5 eCLIP recently performed in K562 cells (https://www.encodeproject.org/experiments/ENCSR238CLX/) the Gemin5 eCLIP bam files were downloaded from www.encodeproject.org and transformed into bed files using bedtools bamtobed. Overlaps between RBS1 and RBS2 peak files and Gemin5 eCLIP reads where calculated using python script fjoin.py (Dataset 2).

MEME software (Multiple EM for Motif Elicitation) (26) was used to identify and characterize the shared motifs in the sequences of peaks including the flanking regions. RNA secondary structure models were predicted using the Mfold software (27). FOLDALIGN software (28) (foldalign.ku.dk/server/index.html) was used to perform pairwise comparisons among the sequences of peaks with flanking regions.

RT-qPCR validation

RNA was isolated from new rounds of CRAFT assays performed with RBS1, RBS2 and TAP polypeptides. Total RNA isolated from cell lysates, without TAP purification, was used as input. Equal amounts of RNAs were used to synthesize cDNA using SuperScript III (Thermo-Fisher) and hexanucleotide mix (Sigma) as primer for the reverse transcription reaction. Primers for quantitative polymerase chain reaction (qPCR) were designed (Primer3 software, http://bioinfo.ut.ee/primer3-0.4.0/primer3/) and tested for amplification efficiency. qPCR was carried out using GoTaq qPCR Master Mix (Promega) according to the manufacturer's instructions on an ABI PRISM 7900HT Fast Real-time PCR system (Applied Biosystems). The primers used are described in Supplementary Table S2. Values were normalized against two constitutive RNAs (HIST1H2BK and MYO5A), present in the RBS1, RBS2, and TAP samples according to the Illumina sequencing data. The comparative cycle threshold method (29) was used to quantify the results (Dataset 3).

For RT-qPCR analysis of Gemin5 mRNA in HEK293 cells expressing TAP, RBS1 or RBS2 we designed two different pairs of primers targeting the regions 1923–2032 and 2338–2440. To determine the steady-state levels of cap-luc and luc-H12 reporter mRNAs in cells expressing TAP or RBS1, we performed RT-qPCR using primers against luciferase (Supplementary Table S2) (Dataset 3). The values were normalized against the constitutive mRNA encoding RPL11 (30).

RNA decay

HEK293 cells, transfected with TAP or RBS1 constructs and either pCap-luc-EcoRV or pCap-luc-H12 constructs, were treated with actinomycin D (5 μg/ml) (31,32) 20 h post-transfection. Cells were harvested at the indicated times of treatment, the RNA was extracted with TRIzol and subjected to RT-qPCR analysis (Dataset 3). Specific primers for luciferase and RPL11 are described in Supplementary Table S2. The data are represented as the log of remaining transcript, and the half-life was determined by the equation of the adjusted line.

Expression and purification of proteins

Escherichia coli BL21 transformed with plasmids pET-G5$_{\Delta 1365–1394}$ (RBS1) and pET-G5$_{1383–1508}$ (RBS2) growing at 37°C were induced with IsopropylβD-1-thiogalactopyranoside (IPTG) 0.5 mM during 2 h. Bacterial cell lysates were prepared in binding buffer (20 mM NaH_2PO_4, 500 mM NaCl, 20 mM Imidazole) using a French press, and cell debris was eliminated by centrifugation at 16 000 g 30 min at 4°C twice. The clear lysates were loaded in His-GraviTrap columns (HealthCare) and the recombinant proteins were eluted using Imidazole 500 mM. Proteins were dialyzed against phosphate buffer pH 6.8, 1 mM DTT, and stored at −20°C in 50% glycerol (14). The purified proteins were analyzed by SDS-PAGE (Supplementary Figure S2).

RNA electrophoretic mobility shift assay

RNA probes were prepared as described (33). Briefly, transcripts were uniformly labelled using α^{32}P-CTP (500 Ci/mmol), T7 RNA polymerase (10 U), and linearized plasmid (1 μg). RNA was extracted with phenol–chloroform, ethanol precipitated and resuspended in TE to a concentration of 0.04 pmol/μl. RNA integrity and mobility as a single band was examined in 6% acrylamide 7 M urea denaturing gel electrophoresis.

RNA-binding reactions were carried out in 10 μl of RNA-binding buffer (40 mM Tris–HCl pH 7.5, 250 mM NaCl, 0.1% (w/v) βME) for 20 min at 4°C. Increasing amounts of protein were incubated with a constant concentration of ^{32}P-labeled RNA (<2 nM), prepared in a mix sufficient for all points of the curve. Electrophoresis was performed in native 6.0% (29:1) polyacrylamide gels. The gels were run in TBE buffer (90 mM Tris–HCl pH 8.4, 64.6 mM boric acid, 2.5 mM EDTA) at 100 V at 4°C for 45 min. The ^{32}P-labeled RNA and retarded complexes were detected by autoradiography of dried gels. The percentage of the retarded complex was calculated relative to the free probe, run in parallel (Dataset 4). Graphs representing the adjusted curves obtained from the quantification of at least three independent gel-shift assays were represented using GraphPad PRISM 6.01.

Immunodetection

Gemin5 and Xpress-G5 proteins were immunodetected using anti-Gemin5 (Novus) antibody. TAP peptide, RBS1-TAP and RBS2-TAP proteins were detected with anti-CBP (Abcam). Immunodetection of tubulin (Sigma) was used as loading control. Secondary antibodies (Thermo Scientific) were used according to the manufacturer's instructions. Quantification of the signal detected was done in the linear range of the antibodies.

RNA pull-down

HEK293 stable transformant cells expressing the MS2-HB protein were obtained using the calcium phosphate procedure with plasmid pMS2-HB (Addgene plasmid # 3557

(20)). Cells were diluted 24 h later into fresh DMEM containing puromycin (3 μg/ml). Individual puromycin resistant clones grown 1–2 weeks later were expanded. After monitoring the expression of MS2-HB protein, cells were frozen until needed.

Confluent monolayers of stable transformant cells HEK293-MS2-HB (10×10^6) were transfected with the constructs pCAP-luc, pCAP-H12-MS2h and pCAP-H12d-MS2h using lipofectamine (Thermo Scientific) according to the manufacturer's instructions. Monolayers were washed 24 h post-transfection with ice-cold PBS prior to UV irradiation (400 mJ/cm^2 at 254 nm). Lysates were prepared in Lysis Buffer PD [20 mM Tris–HCl pH 7.5, 100 mM KCl, 5 mM $MgCl_2$, 0.5% NP-40, 10 mM DTT, protease inhibitors (Merck) and 0.5 U/μl RNase OUT (Thermo Scientific)]. After incubation on ice 10 min, cell debris was discarded by spinning twice at 16 000 g 10 min 4 °C, and the protein concentration in the supernatant was measured by Bradford assay. Streptavidin-coupled Dynabeads (M280) (Thermo Scientific) (20 μl/sample), washed once with Binding&Washing buffer (5 mM Tris–HCl pH 7.5, 1 M NaCl, 0.5 mM EDTA), and three times with Lysis Buffer PD, were resuspended in 20 μl PBS, prior to incubation with the cell supernatant (500 μg total protein) in a final volume of 100 μl Lysis Buffer PD during 2 h at 4°C in a rotating wheel. Aliquots (1%) were taken at time 0 as input samples. Beads were washed three times with 5 volumes of Lysis Buffer PD, 5 min at 4°C. Finally, beads were boiled in SDS-loading buffer and the eluted proteins were resolved by SDS-PAGE. WB analysis was performed using anti-Gemin5. Independent pull-down assays were conducted three times.

Gemin5 polypeptides expression, siRNA interference and luciferase activity assays

HEK293 cell monolayers (2×10^5) were cotransfected with a plasmid expressing luciferase, with or without H12, H34 or H12d sequences (pCAP-luc-EcoRV, pCAP-luc, pCAP-H12-MS2h, pCAP-luc-H12, pCAP-luc-H34 or pCAP-luc-H12d) and a plasmid expressing TAP, RBS1, RBS2 or Gemin5 (pcDNA3-CTAP, pcDNA3-CTAP-RBS1, pcDNA3-CTAP-RBS2 or pcDNA3-Xpress-G5) using lipofectamine (Thermo Scientific).

siRNA targeting Gemin5 mRNA (siRNAG5 CCUUAA UCAAGAAGAGAAAUU) or a control sequence (siRNA-control AUGUAUUGGCCUGUAUUAGUU) were purchased from Dharmacon. HEK293 cells were treated with 100 nM siRNA using lipofectamine 2000 (Thermo Scientific), 30 h prior to transfection of plasmids pCAP-luc-EcoRV or pCAP-luc-H12.

Cell lysates were prepared 24 h post-transfection in 100 μl lysis buffer (50 mM Tris–HCl pH 7.8, 100 mM NaCl, 0.5% NP40). The protein concentration in the lysate was determined by Bradford assay. Equal amounts of protein were loaded in SDS-PAGE to determine the efficiency of interference, as well as the expression of Gemin5-Xpress, RBS1 and RBS2 polypeptides. Luciferase activity (RLU) was normalized to the amount of protein (Dataset 5). Each experiment was repeated independently at least three times. Values represent the mean ± SD.

Statistical analyses

Statistical analyses for experimental data, including RT-qPCR, luciferase activity, and quantification of the band intensity in western blot analysis, were performed as follows. Each experiment was repeated independently at least three times. Values represent the estimated mean ± standard deviation. We computed *P*-values for a difference in distribution between two samples with the unpaired two-tailed Student's *t*-test. Differences were considered significant when $P < 0.05$. The resulting *P*-values were graphically illustrated in figures with asterisks as described in figure legends.

RESULTS

The non-canonical RNA-binding domains of Gemin5 recognize distinct RNAs

We have previously described the presence of a bipartite non-canonical RNA-binding site on the C-terminal region of Gemin5 composed of the domains RBS1 and RBS2 (14). These domains show different behavior concerning IRES-dependent translation, and also distinct capacity to interact with RNA. These results prompted us to identify cellular RNA targets of these domains.

The development of crosslinking and immunoprecipitation (CLIP)-based methodologies to study RNA-protein interactions has represented a breakthrough in understanding the complexity and dynamics of mRNPs (34–38). Here, we set up a crosslink-RNA affinity-tandem (CRAFT) procedure to isolate RNAs crosslinked to TAP-tagged RBS1 and RBS2 polypeptides (Figure 1A), and the TAP peptide alone as a control of unspecific binding. After purification of the UV-crosslinked ribonucleoprotein complexes, the RNAs specifically associated to each of these polypeptides were converted to complementary DNA after adaptor ligations, and subsequently identified by high-throughput sequencing using an Illumina platform (Figure 1B). The RNA-binding properties of Gemin5 indicated that recognition of different RNA targets strongly depends upon RNA structure (10,15,17). Thus, to obtain information on the recognition of the entire target, which presumably contains a folded RNA region longer than the protein-binding motif usually determined in conventional CLIP procedures (21,22), RNase treatment step was omitted from the UV-crosslinked complexes purification. This procedure rendered about 3×10^6 reads for RBS1 and for RBS2 (Supplementary Figure S1A). Then, the combined compiled reads obtained with the control TAP replicates, were subtracted from the overlaps of the RBS1 or RBS2 replicates mapped to the human genome (GRCh37/hg19). This stringent procedure, chosen to eliminate unspecific targets, finally rendered 308 and 298 unique RNA hits associated to RBS1 and RBS2, respectively (Dataset 1). Comparison of the compiled reads identified on each replicate (Supplementary Figure S1B) yielded a correlation coefficient of 0.85 for RBS1, and 0.69 for RBS2.

Computational analysis of the RBS1 and RBS2 compiled reads showed that the hits of RBS1 and RBS2 spanned 22 bases in average, mostly occurring as single hit/gene for each RBS domain (Figure 1C). RBS1 targets, however, had a few cases with three and four hits within the same gene. This result suggested the presence of a relatively long folded

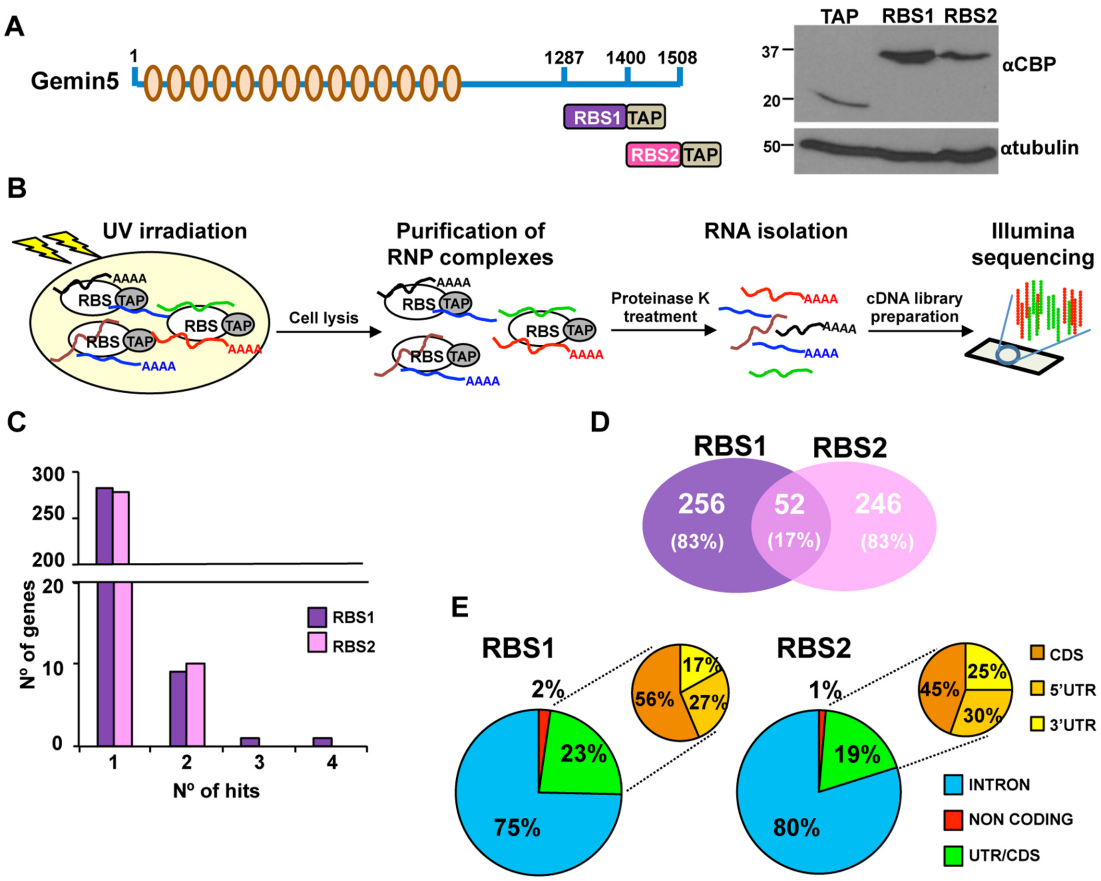


Figure 1. Identification of cellular RNAs associated with the non-canonical RNA-binding domains of Gemin5. (**A**) Schematic of Gemin5 protein. Light green ovals depict the WD40 motifs located within the N-terminal domain. Numbers indicate the amino acid residues flanking TAP-tagged RBS1 and RBS2 polypeptides. Expression of the TAP-tagged RBS1, RBS2 and the control TAP polypeptide in transfected HEK293 cell lysates, monitored by WB using anti-CBP (recognizing the TAP peptide). Tubulin was used as a loading control. (**B**) Overview of CLIP procedure. Lysates obtained from UV-irradiated cells expressing the TAP-tagged RNA-binding site (RBS) polypeptides were subjected to TAP purification to obtain the ribonucleocomplexes bound to the RBS polypeptides, followed by proteinase K treatment to disrupt the complexes, and isolation of bound RNAs. After cDNA libraries preparation, sequences were identified by high-throughput Illumina sequencing. (**C**) Histogram depicting the number of hits per gene. (**D**) Number (and percentage) of overlapping and specific hits associated to RBS1 or RBS2 domains. (**E**) Pie chart depicting the percentage of hits associated to RBS1 or RBS2 domain belonging to distinct RNA types.

RNA sequence, which led to the identification of various hits within the same RNA as a consequence of the library preparation. Overlapping hits between RBS1 and RBS2 were limited to 17%. Hence, most RNA targets (83%) were exclusive of each domain (Figure 1D), suggesting that these domains assemble distinct ribonucleprotein complexes. Of note, the targets of both RBS1 and RBS2 were identified in mRNA sequences (UTR and CDS) (Figure 1E). The high percentage of intronic sequences (75%) was surprising. However, comparison to the intronic sequences in Gemin5 eCLIP data deposited in ENCODE (Supplementary Figure S3) revealed very similar results (89.54%), regardless of the independent approach and differences in cell line. The percentage of hits mapped to noncoding RNAs was scarce both in RBS1 and RBS2 (1 and 2%, respectively). Importantly, no snRNAs were associated to any of these domains of Gemin5. This result unequivocally confirms the presence of different RNA-binding regions in the multifunctional Gemin5 protein, one involved in recognition of snRNAs and cap-binding located at the WD repeats within the N-terminal moiety (11–13), and another one, the non-canonical RBS located at the C-terminal moiety involved in IRES recognition (14). Moreover, the RNA targets of RBS1 and RBS2 differ among themselves (Figure 1D and E), a further evidence for two distinct domains on the C-terminal region of the protein.

The RNA targets of RBS1 domain are enriched in GC-content

Computational search of conserved primary sequence motifs on the targets recognized by RBS1 or RBS2 polypeptides using MEME software (26) revealed differences among these domains. RBS1 targets yielded one GC-rich motif, 29 nts width. RBS2 targets were enriched in two different motifs; the most frequent was GC-rich, about 47 nts width, and the less abundant was AU-rich, 29 nts width (Figure 2A). According to *in silico* RNA structure prediction (27) about 80% of the hits included in the RBS1 motif were predicted to fold as a thermodynamically stable structure ($\Delta G < -20$ kcal /mol) (Supplementary Figure S4). A similar value was obtained for the GC-rich RBS2 hits, while

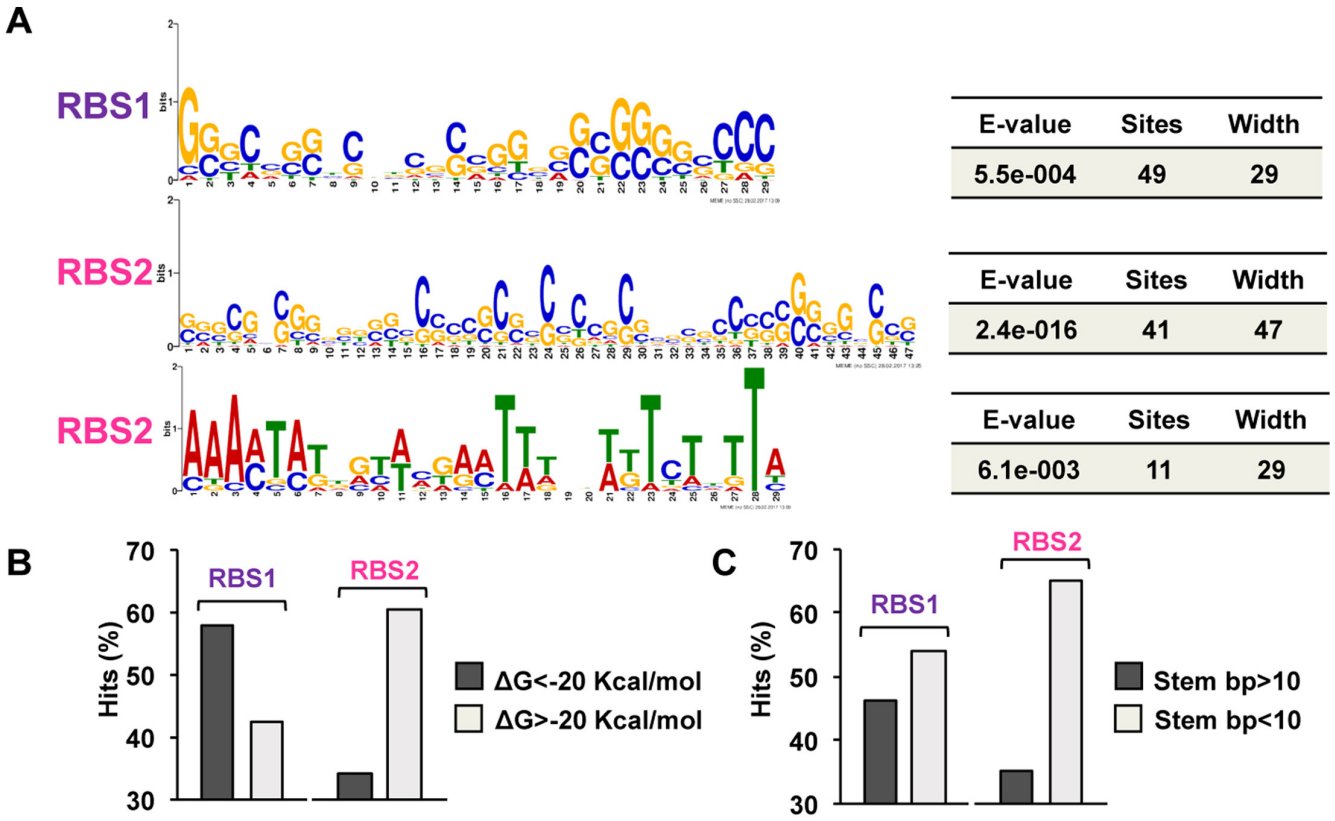


Figure 2. Computational structural analysis of RNA targets of RBS1 and RBS2 domains. (**A**) Sequence consensus logo predicted by MEME for the RNA targets of RBS1 and RBS2. Statistical significance (E-value), number of sequences (sites) and number of nucleotides (width) of the motif are depicted on the right. Histograms depicting the percentage of hits with predicted stable structure ($\Delta G < -20$ kcal/mol) (**B**) or a long stem (bp > 10) (**C**) considering the most abundant hits (25 hits for RBS1 and 20 for RBS2).

none of the AU-rich hits were predicted to have thermodynamically stable secondary structures (Supplementary Figure S4).

Computational analysis focused on the most abundant hits indicated that 60% of the RBS1 hits could fold as a stable hairpin ($\Delta G < -20$ kcal/mol). In contrast, only 35% of the RBS2 most abundant targets had a similar predicted free energy (Figure 2B). Confirming these data case-by-case, 46% of the RBS1 top targets were predicted to adopt a stem longer than 10 bp, as opposed to RBS2, for which only 35% were predicted to fold as a long hairpin (Figure 2C). Interestingly, FOLDALIGN comparison (28) of the predicted structure of individual RBS1 top targets with the most abundant hit of RBS1 revealed a strong similarity (60%) (Supplementary Table S3). The presence of an internal long stem seems to be a feature of the RBS1 most abundant hits. Taken together, these data suggested that RBS1 recognition of the targets depends on RNA structure.

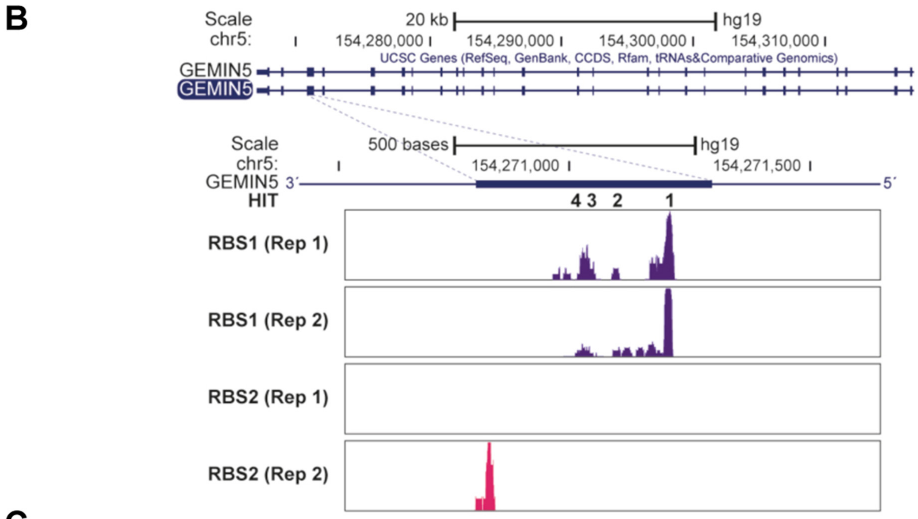
The RBS1 domain interacts directly with the Gemin5 mRNA

To our surprise, the most abundant RBS1 hit was a sequence belonging to Gemin5 mRNA. Noteworthy, four hits were recognized within the mRNA encompassing about 400 nts (Figure 3A), located in the RBS1 coding region. None of these hits were identified by RBS2 (Figure 3B), indicating that these sequences are exclusively recognized by RBS1 polypeptide. To confirm that the Gemin5 mRNA was a bona-fide target of RBS1, we designed specific primers hybridizing inside this region to perform RT-qPCR using RNA samples obtained from new rounds of CRAFT. Gemin5 mRNA was strongly enriched in RBS1-associated RNAs relative to the control TAP-RNA sample, as well as relative to the input-RNA sample (Figure 3C). Furthermore, we did not detect the Gemin5 mRNA in the RBS2-associated RNAs above the TAP-RNA or the input-RNA samples (Figure 3C), confirming that Gemin5 mRNA is an RBS1 specific ligand. Validation of additional targets of RBS1 (SLC12A5 and GOLGA8A) and RBS2 (MYO9A) (Supplementary Figure S5), further confirmed the reliability of this study.

The unexpected result that RBS1 protein recognized Gemin5 mRNA within its own coding region prompted us to define more precisely the RBS1 recognition motif on the mRNA region monitoring RNA-protein complex formation *in vitro* in the absence of other cellular components. To this end, we designed RNA probes covering each of the stronger hits to perform gel shift analysis (Figure 4A). Since the four hits identified in Gemin5 mRNA lie close to each other, we considered the possibility that sequences around the hit could contribute to the interaction of RBS1 with its target. Therefore, we generated probe H12 containing hits 1 and 2, and their interspace sequence. Likewise, probe H34 contained hits 3 and 4, and the short interspace se-

A

HIT	READS	CHROMOSOME	START POSITION	END POSITION	ENSEMBL	GENE	TYPE
1	14	chr5	154271122	154271206	ENSG00000082516	GEMIN5	CDS
2	3	chr5	154271028	154271055	ENSG00000082516	GEMIN5	CDS
3	7	chr5	154270932	154270999	ENSG00000082516	GEMIN5	CDS
4	3	chr5	154270870	154270923	ENSG00000082516	GEMIN5	CDS



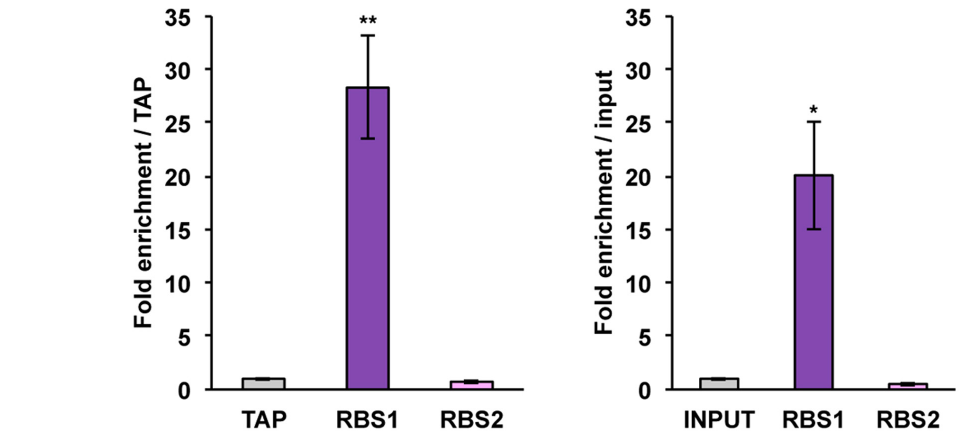


Figure 3. Validation of Gemin5 mRNA as target of the RBS1 domain. (**A**) Location of Gemin5 mRNA hits exclusive for RBS1 domain obtained after filtering the compiled reads with the TAP-peptide associated RNA targets. (**B**) Genome browser screenshot showing Gemin5 gene and the read maps obtained for the replicates of RBS1 and RBS2-associated RNAs. Position of hits 1, 2, 3 and 4 is indicated. (**C**) RT-qPCR fold-enrichment of Gemin5 mRNA relative to TAP (left panel), or relative to input (right panel), in RNA samples obtained from new rounds of CRAFT. Values represent the mean ± SD obtained in three biological independent assays. *P* values were calculated using unpaired Student's *t* test (*$P < 0.05$, **$P < 0.01$).

quence. Then, uniformly labeled probes H12 and H34 were incubated with increasing amounts of purified polypeptide His-RBS1, and the complexes were fractionated in native gels. Quantification of the retarded complex relative to the free probe indicated that RBS1 yielded about 95% retarded complex formation with probe H12 at the highest concentration of protein (500 nM) (Figure 4B). In contrast, probe H34 resulted in ≤ 5% retarded complex with RBS1. Moreover, both H12 and H34 probes yielded ≤ 5% retarded complex formation with purified RBS2 (Supplementary Figure S6A). These results revealed that binding of Gemin5 mRNA to the RBS1 polypeptide occurs via the H12 region (3857-4035 nts).

Consistent with data shown in Figure 2, the probe H12 was predicted to adopt a stable secondary structure (ΔG –52 kcal/mol) (Supplementary Figure S7A). Subdivision of the RNA sequence H12 into H1 and H2 RNAs (Figure 4A) indicated that H1 hit retained the capacity to fold as a stable hairpin (ΔG –30 kcal/mol), similar to that predicted for H12 RNA (Supplementary Figure S7B). However, subsequent gel-shift assays of protein RBS1 with probes H1 or H2 readily indicated that the efficiency of complex forma-

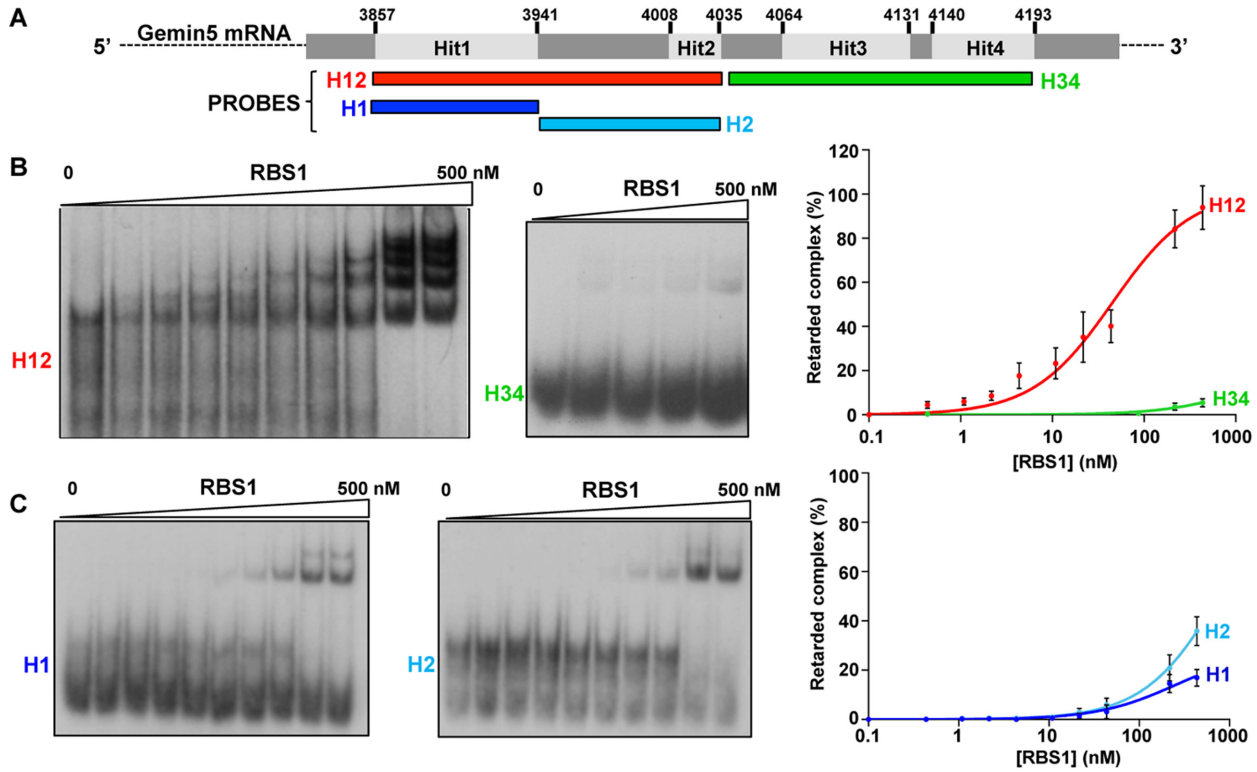


Figure 4. Gemin5 mRNA binding to the RBS1 polypeptide. (**A**) Schematic of the identified hits (1, 2, 3 and 4) on the Gemin5 mRNA. Numbers indicate the nucleotide position referred to the full-length mRNA. The probes (H12, H34, H1 or H2) used in the RNA-protein binding assays are indicated. Gel-shift assay conducted with increasing amounts of purified His-tagged RBS1 and probes H12 or H34 (**B**) and probes H1 or H2 (**C**). Graphs on the right represent the adjusted curves obtained from the quantification of the retarded complex relative to the free probe (mean ± SD) in four independent assays in each case.

tion was lower than H12 (Figure 4C), reaching a percentage of retarded complex formation of ∼20–30%, respectively, at the highest protein concentration (500 nM). As expected from the poor binding to H12, RBS2 did not interact with either H1 or H2 transcripts (Supplementary Figure S6B). To further confirm this result we used an RNA differing in nucleotide sequence but with the capacity to adopt a stem–loop structure. The efficiency of retarded complex formation with RBS1 protein was similar to H2, but significantly lower than H12 (Supplementary Figure S6C). Consistent with previous data, the protein RBS2 did not interact with the control probe.

These results strongly suggest that the RNA-protein complex formation is favored by a sequence that tends to adopt a stable hairpin (≥10 bp), and that the sequence around the hairpin contributes to the binding as well. We conclude that the H12 region of Gemin5 mRNA harbors a sequence and a structure fold that provides the recognition motif for RBS1 polypeptide. Furthermore, since RBS2 polypeptide interacted rather weakly with this probe, we conclude that the RNA-binding domain of the C-terminal region is located within RBS1.

Destabilization of H1 stem reduces the binding to Gemin5 RNA *in vitro* and in living cells

As mentioned above, the RBS1 targets were predicted to fold as secondary structure including an internal long stem–loop. Moreover, the H12 probe containing the H1 stem yielded the highest retarded complex capacity with RBS1 protein (Figure 4B). These results prompted us to investigate the binding capacity of a modified H12 RNA sequence that destabilizes the H1 stem. Hence, we prepared the H12d RNA substituting three tracts of five consecutive nts to adenines within H1, preserving intact the H2 region (Supplementary Figure S8A). This sequence was predicted to adopt a less stable secondary structure (ΔG –32 kcal/mol) than the H12 counterpart (ΔG –52 kcal/mol, Supplementary Figure S7A). Gel-shift assays performed with probe H12d and increasing amounts of protein RBS1 revealed a decreased retarded complex (70%) relative to H12 probe, at the highest concentration of protein (500 nM) (Figure 5A). As expected from results shown for RBS2 with other probes (Supplementary Figure S6) the percentage of retarded complex with RBS2 polypeptide was very low (Supplementary Figure S8B).

Then, to determine whether transcripts H12 and H12d are recognized by Gemin5 protein in the cellular context, we transfected stable transformants HEK293 expressing a biotinylated MS2 protein with plasmids that produce transcripts tagged with MS2 hairpins (Figure 5B). This system allows purification of factors associated to the RNA via pull down of the MS2 protein. Thus, RNA pull-down samples of cell lysates expressing cap-MS2h, H12-MS2h, and H12d-MS2h transcripts were used to immunodetect Gemin5. As shown in Figure 5C, the H12-MS2h RNA

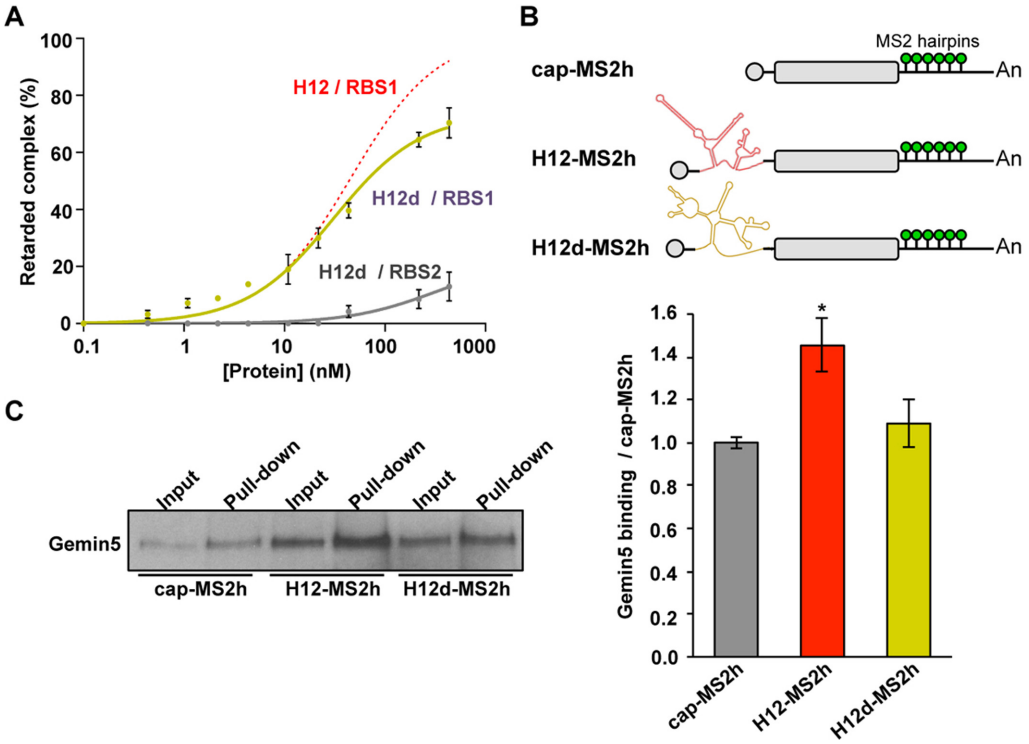


Figure 5. Destabilization of the H1 stem reduced the binding to Gemin5 RNA. (**A**) Graph representing the adjusted curves obtained from the quantification (mean ± SD) of independent gel-shift assays using H12d probe incubated with RBS1 ($n = 3$) or RBS2 ($n = 4$) purified proteins. (**B**) Diagram of the RNAs used to perform RNA pull-downs. (**C**) RNA pull-down assays performed with cell lysates expressing cap-MS2h, H12-MS2h or H12d-MS2h mRNAs. A representative example of the immunoblotting of the endogenous Gemin5 protein is shown on the left panel. Quantification of the pull-down/input intensity of Gemin5 measured in three independent assays with H12 or H12d mRNAs relative to cap-luc mRNA (right panel). Values represent the mean ± SD (*$P < 0.05$ by Student's t test).

efficiently pulled down Gemin5. Additionally, the relative intensity of the pull-down/input obtained for H12-MS2h mRNA was higher than that observed for cap-MS2h or H12d-MS2h mRNAs (Figure 5C, left panel). Quantification of the intensity of Gemin5 bands of three independent assays revealed a statistically significant enrichment of Gemin5 (1.5-fold) in H12-MS2h RNA pull-down relative to the cap-MS2h control transcript. In contrast, no significant enrichment was observed for H12d-MS2h RNA (Figure 5C, right panel).

Therefore, we conclude that destabilization of the stem H1 in the context of H12 RNA reduces the binding of RBS1 protein *in vitro,* and also decreases the interaction with the endogenous Gemin5 protein in the cell environment.

The RBS1 domain stimulates translation of the endogenous Gemin5 mRNA

The results shown above strongly suggested a crosstalk between the RBS1 domain of Gemin5 protein and the H12 region of Gemin5 mRNA. Thus, we sought to investigate whether RBS1 had any effect on the endogenous Gemin5 mRNA expression, determined by immunostaining of the full-length Gemin5 protein. Interestingly, a 2.1-fold increase of the intensity of the immunodetected Gemin5 protein was observed in cells expressing RBS1, relative to control cells (Figure 6A). This effect was not observed in cells expressing the RBS2 polypeptide. Next, we tested if the increased levels of Gemin5 protein in cells expressing RBS1 could be due to altered amounts of endogenous mRNA. The relative amounts of Gemin5 mRNA determined using two different pairs of primers for the RT-qPCR were not statistically significant in cells expressing RBS1 or RBS2, showing that expression of RBS1 in HEK293 cells does not increase the steady-state content of Gemin5 mRNA (Figure 6B).

Considering these data, we conclude that the RBS1 domain of Gemin5 exerts an autoregulatory role on Gemin5 mRNA expression, which could not be attributed to increased levels of endogenous Gemin5 mRNA. Rather, we envisioned that it could occur at the translation step.

The role of RBS1 domain on translation depends upon the mRNA target

The results shown above suggested to us that the interaction of RBS1 domain of Gemin5 with its own mRNA could have profound functional implications, possibly controlling its own translation. To further address this possibility we measured the effect of the RBS1 polypeptide expression on translation control, relative to the RBS2 polypeptide and the full-length Gemin5 protein, using three reporter transcripts. The cap-luc transcript lacking any Gemin5 mRNA sequences served as control; the luc-H12 and luc-H34 transcripts contain the sequences H12 or H34 at the 3′end of the reporter mRNA, respectively (Figure 7A). Hence, luciferase activity expressed from the cap-luc transcript reflected the

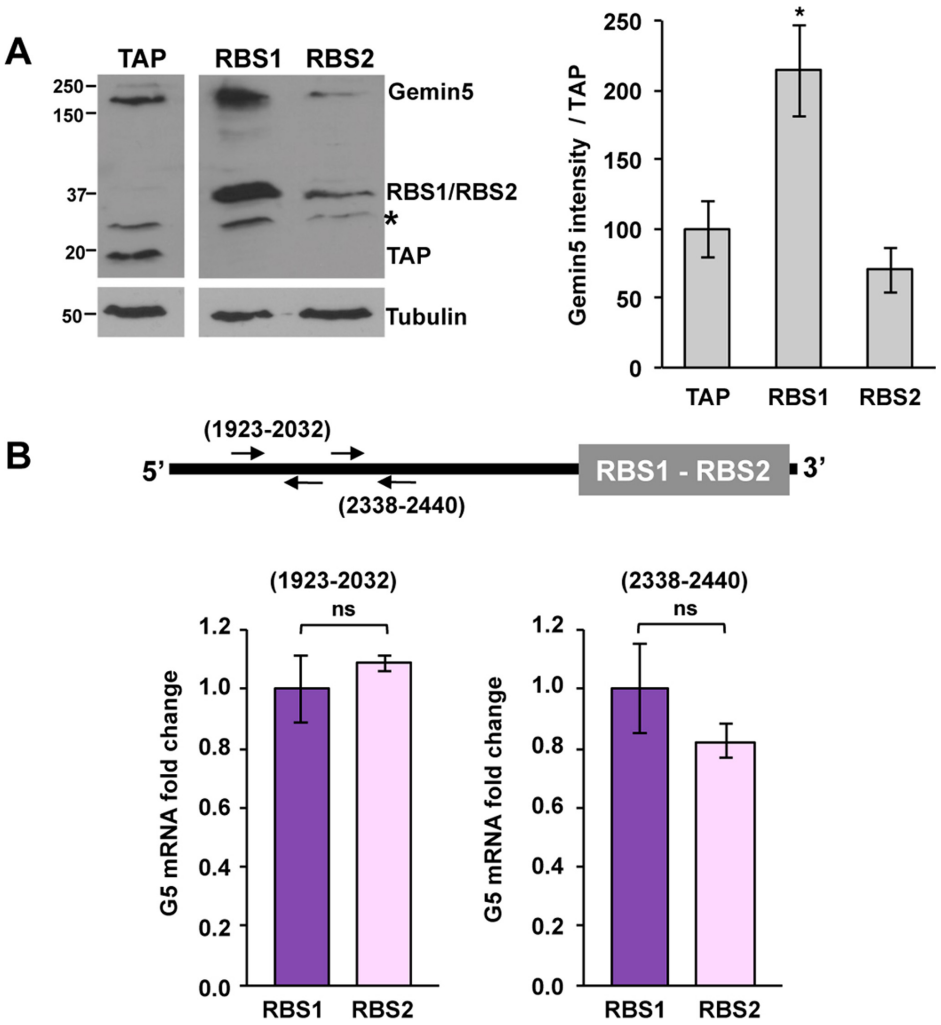


Figure 6. RBS1 domain stimulates translation of the endogenous Gemin5 mRNA. (**A**) Western blot analysis of endogenous Gemin5 in cells expressing TAP, RBS1 or RBS2 polypeptides. Gemin5 was immunodetected with anti-Gemin5 antibody, while TAP, RBS1, and RBS2 polypeptides were detected with anti-CBP antibody. Tubulin was used as loading control. Asterisk depicts an unspecific product. Quantification of the intensity of Gemin5 bands from two independent assays in cells expressing RBS1 or RBS2 proteins relative to cells expressing TAP polypeptide (loaded on the same blot in separate lanes). Values represent the mean ± SD, (*$P < 0.05$ by Student's t test). (**B**) RT-qPCR of Gemin5 mRNA in cell extracts expressing RBS1 or RBS2 polypeptides. The aproximate location of two different pairs of primers, amplifying the regions 1923–2032 and 2338–2440 of Gemin5 mRNA are shown on the top panel. These primers hybridize upstream of the sequence encoded in plasmids expressing RBS1 or RBS2. RNA fold change was normalized to TAP, and then to RBS1.

potential effect of the RBS1, RBS2 and Gemin5 proteins on cap-dependent translation, whereas luciferase activity from the transcripts luc-H12 or luc-H34 indicated additional effects of these proteins dependent upon the presence of the RBS1-target sequences on the reporter mRNA. Expression of the TAP, RBS1, RBS2 and Xpress-Gemin5 (18) proteins was assessed 24 h post-transfection of HEK293 cells by immunobloting. Similar levels of expression were observed in all cases, irrespectively of the coexpression of these proteins with any of the transcripts, cap-luc, luc-H12 or luc-H34 (Figure 7B).

The effect of RBS1 or RBS2 polypeptides and Gemin5 protein expression on luciferase activity from cap-luc transcript normalized to cells expressing the TAP peptide control is shown in Figure 7C (left panel). The RBS1 protein induced a statistically significant inhibition of luciferase translation, reaching values near 50% of the TAP control. In contrast, RBS2 expression barely reduced luciferase activity (80%). Notably, a decrease of luciferase activity similar to that of RBS1-expressing cells was observed in cells that express Xpress-G5 (Figure 7C, left panel). The Gemin5 inhibitory effect on translation is in agreement with previous data using different RNA reporters (16,18).

Next, we analyzed the consequences of having the sequence H12 inserted into the 3′UTR of the reporter mRNA in cells expressing RBS1, RBS2 or Xpress-G5 relative to cells expressing the control TAP peptide. We noticed that the negative effect of RBS1 and Xpress-G5 proteins in luciferase activity from cap-luc transcript was counterbalanced by the presence of the H12 sequence on the mRNA (Figure 7C, right panel). No changes were observed in cells expressing RBS2. Compared to H12, the H34 sequence partially recovered the negative effect when RBS1 or Gemin5 were co-expressed (Supplementary Figure S9A).

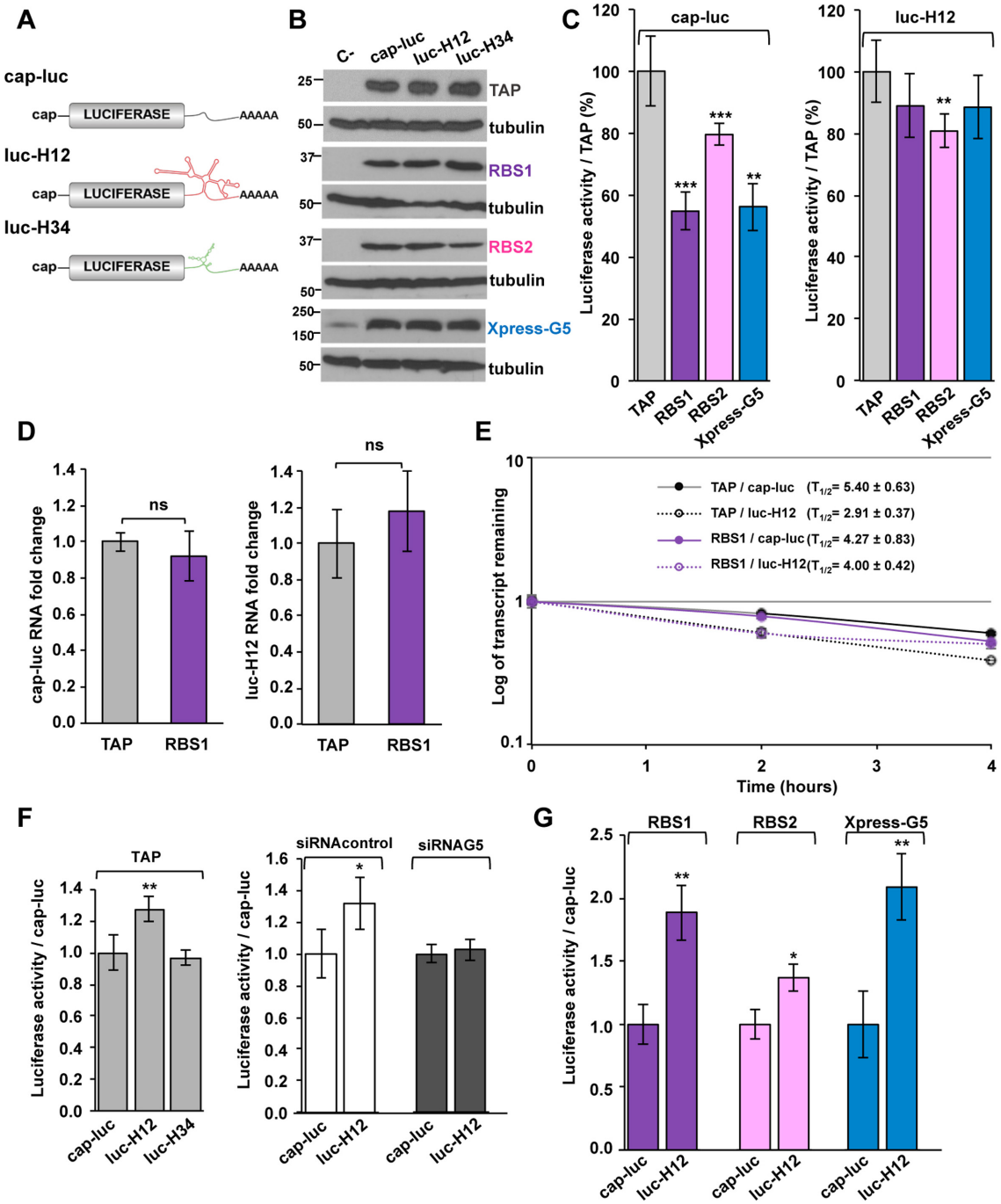


Figure 7. Effect of H12 sequence on mRNA translation. (**A**) Diagram of transcripts cap-luc, luc-H12 and luc-H34. (**B**) Expression of TAP, RBS1, RBS2 and Xpress-G5 proteins in HEK293 cells co-transfected with plasmids expressing cap-luc, luc-H12, or luc-H34 RNAs. Lane C- depicts mock-transfected cells. TAP, RBS1 and RBS2 were immunodetected with anti-CBP, Xpress-G5 with anti-Gemin5. Tubulin was used as loading control. (**C**) Luciferase activity measured in cell lysates expressing cap-luc mRNA (left panel) or luc-H12 mRNA (right panel) cotransfected with RBS1, RBS2, or Xpress-G5. Values are normalized to cells expressing the control TAP peptide ($n = 5$). (**D**) RBS1 expression does not affect steady-state RNA levels of luciferase reporter mRNAs. RT-qPCR analysis of cap-luc and luc-H12 mRNAs in HEK293 cells expressing TAP or RBS1. Values represent the mean ± SD obtained in three independent assays (ns, not statistically significant). (**E**) Half-life of luciferase reporter mRNAs. HEK293 cells, transfected with TAP or RBS1 constructs, were treated with actinomycin D (5 μg/ml) 20 h post-transfection. Cells were harvested at different times post-treatment; the total RNA was extracted and subjected to RT-qPCR. (**F**) Normalized luciferase activity obtained for luc-H12 and luc-H34 transcripts relative to cap-luc mRNA in cells expressing TAP (left panel, $n = 11$). Normalized luciferase activity in control and Gemin5-depleted cells for luc-H12 transcript, relative to cap-luc mRNA (right panel, $n = 3$). (**G**) Normalized luciferase activity obtained for luc-H12 transcript relative to cap-luc mRNA in cells expressing RBS1, RBS2, or Xpress-G5 proteins ($n = 5$). Values represent the mean ± SD (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$ by Student's t test).

To determine if the increase in luciferase activity could be due to different content of reporter luc-H12 RNA, we analyzed the levels and the half-life of cap-luc and luc-H12 RNAs in cells expressing the control TAP or RBS1 polypeptides. No significant difference in RNA fold change was observed (Figure 7D). In addition, no major differences were observed in the half-life of cap-luc or luc-H12 RNAs in cells expressing TAP or RBS1 (Figure 7E). Hence, RBS1 expression does not affect the steady-state RNA levels or the half-life of luciferase reporter mRNAs.

Taken together, the results observed in cells expressing the RBS1 polypeptide allowed us to conclude that the negative effect of Gemin5 on translation concurs with the RBS1 domain, and more importantly, that the presence of H12 sequence of the Gemin5 mRNA counteracts this effect.

To analyze further the effect of H12 and H34 sequences on translation, independently of the inhibitory effect of the RBS1 or Xpress-G5 proteins, we normalized the levels of luciferase activity produced from constructs luc-H12 and luc-H34 relative to the levels produced from cap-luc in control cells expressing TAP (Figure 7F, left panel). The activity of luc-H12 construct was higher (1.3-fold) than cap-luc, while no changes were observed with the luc-H34 transcript. Then, to determine if the stimulatory effect of luc-H12 transcript observed in control cells could depend on the endogenous Gemin5 protein, we monitored luciferase expression from luc-H12 construct in cells depleted of Gemin5. Consistent with earlier reports of downregulatory effects of Gemin5 on cap-dependent translation (16,18), silencing of Gemin5 led to an increase of luciferase activity from the cap-luc construct (Supplementary Figure S10). Importantly, the increased activity of luc-H12 transcript in siRNAcontrol-transfected cells was reverted when the endogenous Gemin5 was silenced (siRNAG5) (Figure 7F, right panel). These results indicated that presence of the H12 sequence in a given mRNA is sufficient to stimulate its translation, and that the Gemin5 protein is necessary for this stimulatory effect.

The next question was whether the H12 sequence stimulation of translation depends upon RBS1, RBS2 or Gemin5 proteins expression. In cells expressing RBS1, translation of luc-H12 construct was higher than translation of cap-luc (1.9-fold higher) (Figure 7G). In contrast, the effect of sequence H12 in cells expressing RBS2 polypeptide remained similar to the cells expressing the control TAP polypeptide (compared to Figure 7F, left panel). Importantly, translation of luc-H12 in cells expressing the full length Xpress-G5 showed a stimulatory effect (2.1-fold higher) (Figure 7G), similar to that observed in cells expressing RBS1. A modest effect (1.3-fold) was observed with construct luc-H34 in cells expressing RBS1 or Xpress-G5 (Supplementary Figure S9B). These results indicated that the stimulatory effect on translation induced by H12 sequence is enhanced when RBS1 or Gemin5 proteins are co-expressed with the H12-containing mRNA. Altogether, we conclude that recognition of the H12 RNA sequence by the RBS1 domain of Gemin5 stimulates protein synthesis, hence autoregulating its own mRNA translation.

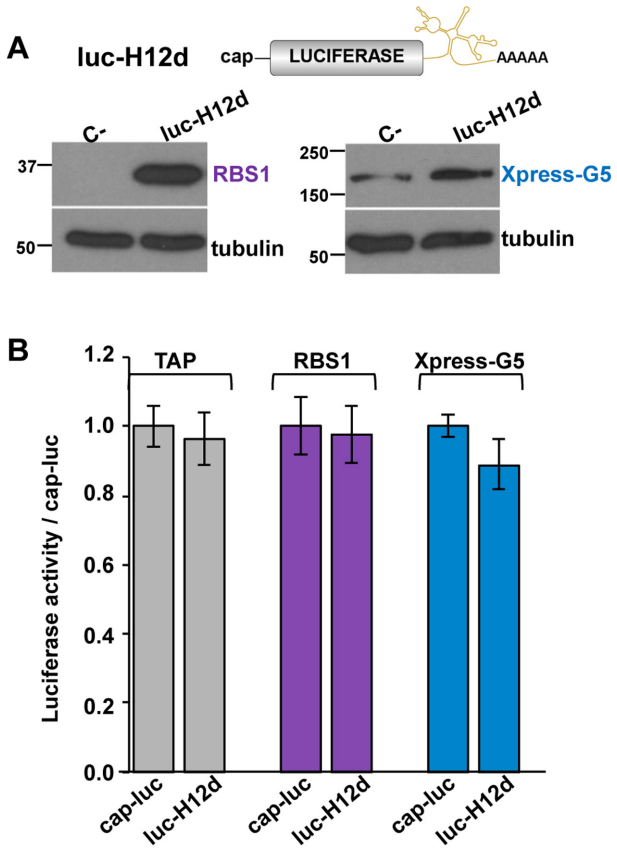


Figure 8. Effect on translation and RNA-binding capacity of a destabilized H12 RNA. (**A**) Diagram of luc-H12d transcript. Expression of RBS1 and Xpress-G5 proteins in HEK293 cells co-transfected with the plasmid that produces the transcript luc-H12d. Lane C- depicts mock-transfected cells. RBS1 was immunodetected with anti-CBP antibody, Xpress-G5 with anti-Gemin5 antibody; tubulin was used as loading control. (**B**) Normalized luciferase activity detected with luc-H12d transcript relative to cap-luc mRNA in cells expressing TAP, RBS1, or Xpress-G5 proteins. Values represent the mean ± SD obtained in three independent assays.

Destabilization of H1 stem impairs the H12 stimulatory effect on translation

Given that destabilization of H1 hairpin induced a decreased binding to Gemin5 *in vivo,* as well as to RBS1 *in vitro* (Figure 5), we analyzed the luciferase activity expressed from a transcript with the sequence H12d inserted on 3′UTR relative to cap-luc in cells expressing RBS1 or Xpress-G5 proteins (Figure 8A). The construct luc-H12d produced similar luciferase activity than the cap-luc transcript, irrespectively of the coexpression of TAP, RBS1 polypeptides or Xpress-G5 protein (Figure 8B). These results demonstrate that the H1 stem–loop plays a central role on the stimulatory effect on translation.

In summary, the results shown here allowed us to propose a model for the role of Gemin5 in translation control (Figure 9). The N-terminal region of the protein harbors a tandem WD40 domain, which recognizes RNA oligonucleotides containing the sm-site and m^7G cap moieties independently (11,13). This region of Gemin5 is also involved in binding to m^7GTP resins (39), and in ribosome binding (16,18), controlling global translation. On the other hand,

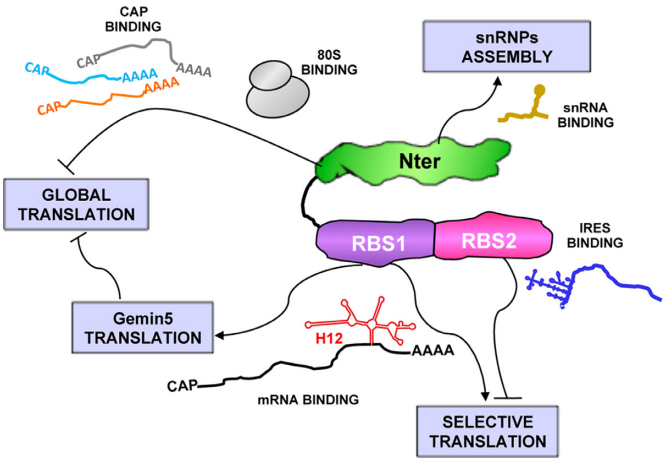


Figure 9. Model for the role of Gemin5 on translation control. Specific domains of the protein recognize distinct RNA targets. The functional domains Nter, RBS1 and RBS2 are colored in green, purple and pink, respectively. WD motifs in the Nter region are responsible for the delivery of snRNAs to the SMN complex (9), governing snRNPs assembly. In addition, G5-Nter interacts with ribosomes via the 60S subunit (16), impacting on global mRNA translation; interaction with the m^7G cap also occurs through this region (11–13). The RBS2 domain modulates IRES-dependent activity (14). The RBS1 domain can perform a dual role in translation; recognition of stable stem–loops by RBS1 stimulates mRNA translation. However, RBS1 represses translation by increasing the levels of Gemin5 protein through RBS1 binding to the H12 region of Gemin5 mRNA.

the C-terminal region harbors a bipartite RNA-binding domain, RBS1 and RBS2 (14), involved in IRES-dependent translation (14,40). Among the cellular targets of RBS1 domain identified in this study, we have found that one of the most abundant hits was Gemin5 mRNA, and that this interaction is mediated by a sequence predicted to fold as a stable stem–loop. Notwithstanding, interaction of RBS1 with the H12 mRNA region counteracts the negative effect of Gemin5 on translation. Thus, the RBS1 domain of Gemin5 can modulate translation in different manners depending on the mRNA target, upregulating H12-like counterparts, or downregulating translation of conventional mRNAs. In the Gemin5 mRNA scenario, we propose that RBS1 domain regulates the availability of this protein to play its multiple roles in gene expression control.

DISCUSSION

Gemin5 is a predominantly cytoplasmic protein that performs critical functions in evolutionary distant organisms. In humans, the highest expression of Gemin5 occurs in the gonads (41,42). In *Drosophila,* loss of Gemin5/Rigor mortis protein is lethal at the larva stage (43). Beyond its capacity to deliver the snRNAs into the SMN complex (10), this multifunctional protein downregulates translation (16,18). These features rely on the capacity of Gemin5 to recognize multiple partners through different functional domains. The RNA targets of Gemin5 include structural motifs, such as the snRNAs recognized by the WD repeat motifs at the N-terminal end (9), and the IRES element recognized by the non-canonical RNA-binding site at the C-terminal end (15). The latter consists of two domains, RBS1 and RBS2, which differ in their capacity to bind RNA, and also in their ability to modulate IRES-dependent translation (14).

Based on these differences, we sought to investigate the cellular RNA targets of RBS1 and RBS2 polypeptides. Here we show that most hits (83%) are exclusive of one domain, reinforcing the idea that the RBS domains of Gemin5 could exert different functions depending on the target RNA. Two observations indicated that not only the sequence but also the RNA conformation could contribute to RNA recognition by RBS1: (i) the most abundant hits consisted of G:C-rich sequences predicted to fold as a stem longer than 10 bp, (ii) four nearby hits mapped within a region of ∼400 nts. These results strongly suggest that a key feature of the RBS1 RNA hits was the ability to fold as a stable hairpin, in agreement with the RNA-binding properties of the C-terminal domain of Gemin5 (15).

Remarkably, the most abundant and exclusive hit of RBS1 polypeptide was the Gemin5 mRNA. In support of our results, data obtained with the K562 cell line deposited in ENCODE (https://www.encodeproject.org/experiments/ENCSR238CLX/) shows 30% overlap between the Gemin5 associated RNAs with our data, in spite of using independent approaches in a different cell line (Dataset 2). More importantly, there is full coincidence in the four peaks observed for Gemin5 recognition of its own mRNA. After validation of the specificity of the RBS1-Gemin5 RNA interaction by RT-qPCR, we narrowed the RBS1 recognition motif to the H12 region (nts 3857–4035) of Gemin5 mRNA performing gel-shift analysis with purified components. In addition, the H12 region recruits the endogenous Gemin5 protein in the cell environment. Notwithstanding, assays carried out with H12d RNA, which destabilizes the internal stem H1, showed a reduced binding to Gemin5 protein in cell lysates, and also to the RBS1 polypeptide *in vitro.* Therefore, H1 seems to provide the core recognition motif within H12, although the sequence flanking H1 also contributes to the binding. For instance, the probe H12 yielded over 95% of the RNA associated in the retarded complex, while formation of retarded complex with H1 probe was drastically reduced (20%). These data suggests that the H1 flanking sequence enhances RBS1 binding, presumably by stabilizing the optimal structure required for protein recognition.

Gemin5 could contribute to translation control in different manners by using distinct functional domains. The negative role in global translation of the N-terminal region of Gemin5 has been related to the capacity to bind the ribosome, and to sediment with heavy polysomes (18). Accordingly, the WD40 domains located in this region recognize the cap structure of RNAs (11–13), consistent with previous reports (39). However, the negative effect of Gemin5 on global translation (16) was also observed following expression of the RBS1 domain in HEK293 cells (Figure 7C). Given that the C-terminal domain of Gemin5 does not interact with the ribosome and does not sediment with heavy polysomes (18), we propose that the negative effect of RBS1 polypeptide on translation is due to the enhancing effect of RBS1 domain on the endogenous Gemin5 mRNA (Figure 7F). Consequently, an increase in the amount of Gemin5 protein could downregulate global protein synthesis (Figure 9).

The unexpected interaction of RBS1 polypeptide with the Gemin5 mRNA led us to further investigate whether this could be relevant to control its own gene expression. Interestingly, the RBS1 domain stimulates expression of the endogenous Gemin5 protein, strongly suggesting that the interaction of RBS1 with the Gemin5 mRNA has a relevant role. Experiments with RNA reporters confirmed that the presence of the H12 sequence on a given RNA stimulates its translation, independently of the position on the mRNA (Supplementary Figure S11). In further support of this conclusion, destabilization of the H12 stem–loop resulted in the loss of the stimulatory effect on translation. These results provide support for the relevance of RNA structure on the RBS1-dependent translation control, and demonstrate a regulatory effect that depends on both, the mRNA target sequence and the expression of the RBS1 domain (Figure 9).

In summary, our data provides a mechanistic basis for the autoregulatory effect of the H12 sequence present on Gemin5 mRNA in conjunction with the expression of the RBS1 domain. Besides, it also suggests that the C-terminal region of Gemin5 plays a multiple role inside the cell. We propose that the RBS1 domain primarily determines RNA-interaction. The RBS2 domain appears to control selective translation, including IRES-driven protein synthesis (14). However, the RBS1 domain downregulates global translation of conventional mRNAs, whereas positively enhances translation of RNA targets carrying stable secondary structure motifs, as shown here for the H12 stem–loop present in Gemin5 mRNA.

DATA AVAILABILITY

The accession number for the FASTQ files from the CRAFT experiments reported in this paper is GEO GSE102268.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank the CRG sequencing unit for sequencing the CRAFT RNA samples, E. Eyras for sharing with us the pyicoclip software, R. Peiro (CBMSO genomics unit) for helping with the genomic data, and C. Gutierrez, A.M. Embarek and M.I. Gismondi for valuable comments on the manuscript.

FUNDING

MINECO [BFU2014-54564, BIO2015-72716-EXP]; Comunidad de Madrid [B2017/BMD-3770]; Institutional grant from Fundación Ramón Areces. Funding for open access charge: MINECO [BIO2015-72716-EXP].
Conflict of interest statement. None declared.

REFERENCES

1. Gehring,N.H., Wahle,E. and Fischer,U. (2017) Deciphering the mRNP code: RNA-bound determinants of post-transcriptional gene regulation. *Trends Biochem. Sci.*, **42**, 369–382.
2. Lunde,B.M., Moore,C. and Varani,G. (2007) RNA-binding proteins: modular design for efficient function. *Nat. Rev. Mol. Cell. Biol.*, **8**, 479–490.
3. Castello,A., Fischer,B., Hentze,M.W. and Preiss,T. (2013) RNA-binding proteins in Mendelian disease. *Trends Genet.*., **29**, 318–327.
4. Otter,S., Grimmler,M., Neuenkirchen,N., Chari,A., Sickmann,A. and Fischer,U. (2007) A comprehensive interaction map of the human survival of motor neuron (SMN) complex. *J. Biol. Chem.*, **282**, 5825–5833.
5. Kroiss,M., Schultz,J., Wiesner,J., Chari,A., Sickmann,A. and Fischer,U. (2008) Evolution of an RNP assembly system: a minimal SMN complex facilitates formation of UsnRNPs in Drosophila melanogaster. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 10045–10050.
6. Grimm,C., Chari,A., Pelz,J.P., Kuper,J., Kisker,C., Diederichs,K., Stark,H., Schindelin,H. and Fischer,U. (2013) Structural basis of assembly chaperone- mediated snRNP formation. *Mol. Cell*, **49**, 692–703.
7. Meister,G., Buhler,D., Pillai,R., Lottspeich,F. and Fischer,U. (2001) A multiprotein complex mediates the ATP-dependent assembly of spliceosomal U snRNPs. *Nat. Cell Biol.*, **3**, 945–949.
8. Will,C.L. and Luhrmann,R. (2001) Spliceosomal UsnRNP biogenesis, structure and function. *Curr. Opin. Cell Biol.*, **13**, 290–301.
9. Battle,D.J., Lau,C.K., Wan,L., Deng,H., Lotti,F. and Dreyfuss,G. (2006) The Gemin5 protein of the SMN complex identifies snRNAs. *Mol. Cell*, **23**, 273–279.
10. Yong,J., Kasim,M., Bachorik,J.L., Wan,L. and Dreyfuss,G. (2010) Gemin5 delivers snRNA precursors to the SMN complex for snRNP biogenesis. *Mol. Cell*, **38**, 551–562.
11. Xu,C., Ishikawa,H., Izumikawa,K., Li,L., He,H., Nobe,Y., Yamauchi,Y., Shahjee,H.M., Wu,X.H., Yu,Y.T. *et al.* (2016) Structural insights into Gemin5-guided selection of pre-snRNAs for snRNP assembly. *Genes Dev.*, **30**, 2376–2390.
12. Tang,X., Bharath,S.R., Piao,S., Tan,V.Q., Bowler,M.W. and Song,H. (2016) Structural basis for specific recognition of pre-snRNA by Gemin5. *Cell Res.*, **26**, 1353–1356.
13. Jin,W., Wang,Y., Liu,C.P., Yang,N., Jin,M., Cong,Y., Wang,M. and Xu,R.M. (2016) Structural basis for snRNA recognition by the double-WD40 repeat domain of Gemin5. *Genes Dev.*, **30**, 2391–2403.
14. Fernandez-Chamorro,J., Pineiro,D., Gordon,J.M., Ramajo,J., Francisco-Velilla,R., Macias,M.J. and Martinez-Salas,E. (2014) Identification of novel non-canonical RNA-binding sites in Gemin5 involved in internal initiation of translation. *Nucleic Acids Res.*, **42**, 5742–5754.
15. Pineiro,D., Fernandez,N., Ramajo,J. and Martinez-Salas,E. (2013) Gemin5 promotes IRES interaction and translation control through its C-terminal region. *Nucleic Acids Res.*, **41**, 1017–1028.
16. Pacheco,A., Lopez de Quinto,S., Ramajo,J., Fernandez,N. and Martinez-Salas,E. (2009) A novel role for Gemin5 in mRNA translation. *Nucleic Acids Res.*, **37**, 582–590.
17. Workman,E., Kalda,C., Patel,A. and Battle,D.J. (2015) Gemin5 binds to the survival motor neuron mRNA to regulate SMN expression. *J. Biol. Chem.* **5**, 528–544.
18. Francisco-Velilla,R., Fernandez-Chamorro,J., Ramajo,J. and Martinez-Salas,E. (2016) The RNA-binding protein Gemin5 binds directly to the ribosome and regulates global translation. *Nucleic Acids Res.* **44**, 8335–8351.
19. Chen,G.I. and Gingras,A.C. (2007) Affinity-purification mass spectrometry (AP-MS) of serine/threonine phosphatases. *Methods*, **42**, 298–305.
20. Tsai,B.P., Wang,X., Huang,L. and Waterman,M.L. (2011) Quantitative profiling of in vivo-assembled RNA-protein complexes using a novel integrated proteomic approach. *Mol. Cell. Prot.: MCP*, **10**, M110 007385.
21. Ray,D., Kazan,H., Cook,K.B., Weirauch,M.T., Najafabadi,H.S., Li,X., Gueroussov,S., Albu,M., Zheng,H., Yang,A. *et al.* (2013) A compendium of RNA-binding motifs for decoding gene regulation. *Nature*, **499**, 172–177.
22. Sugimoto,Y., Konig,J., Hussain,S., Zupan,B., Curk,T., Frye,M. and Ule,J. (2012) Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions. *Genome Biol.*, **13**, R67.

23. Francisco-Velilla,R., Fernandez-Chamorro,J., Lozano,G., Diaz-Toledano,R. and Martinez-Salas,E. (2015) RNA-protein interaction methods to study viral IRES elements. *Methods*, **91**, 3–12.
24. Althammer,S., Gonzalez-Vallinas,J., Ballare,C., Beato,M. and Eyras,E. (2011) Pyicos: a versatile toolkit for the analysis of high-throughput sequencing data. *Bioinformatics*, **27**, 3333–3340.
25. Richardson,J.E. (2006) fjoin: simple and efficient computation of feature overlaps. *J. Comput. Biol.*, **13**, 1457–1464.
26. Bailey,T.L., Johnson,J., Grant,C.E. and Noble,W.S. (2015) The MEME Suite. *Nucleic Acids Res.*, **43**, W39–W49.
27. Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.
28. Sundfeld,D., Havgaard,J.H., de Melo,A.C. and Gorodkin,J. (2016) Foldalign 2.5: multithreaded implementation for pairwise structural RNA alignment. *Bioinformatics*, **32**, 1238–1240.
29. Schmittgen,T.D. and Livak,K.J. (2008) Analyzing real-time PCR data by the comparative C(T) method. *Nat. Protoc.*, **3**, 1101–1108.
30. Lozano,G., Francisco-Velilla,R. and Martinez-Salas,E. (2018) Ribosome-dependent conformational flexibility changes and RNA dynamics of IRES domains revealed by differential SHAPE. *Sci. Rep.*, **8**, 5545.
31. Bensaude,O. (2011) Inhibiting eukaryotic transcription: which compound to choose? How to evaluate its activity? *Transcription*, **2**, 103–108.
32. Jablonski,J.A. and Caputi,M. (2009) Role of cellular RNA processing factors in human immunodeficiency virus type 1 mRNA metabolism, replication, and infectivity. *J. Virol.*, **83**, 981–992.
33. Galan,A., Lozano,G., Pineiro,D. and Martinez-Salas,E. (2017) G3BP1 interacts directly with the FMDV IRES and negatively regulates translation. *FEBS J.*, **284**, 3202–3217.
34. Lee,A.S., Kranzusch,P.J. and Cate,J.H. (2015) eIF3 targets cell-proliferation messenger RNAs for translational activation or repression. *Nature*, **522**, 111–114.
35. Maslon,M.M., Heras,S.R., Bellora,N., Eyras,E. and Caceres,J.F. (2014) The translational landscape of the splicing factor SRSF1 and its role in mitosis. *Elife*, e02028.
36. Yoon,J.H., De,S., Srikantan,S., Abdelmohsen,K., Grammatikakis,I., Kim,J., Kim,K.M., Noh,J.H., White,E.J., Martindale,J.L. *et al.* (2014) PAR-CLIP analysis uncovers AUF1 impact on target RNA fate and genome integrity. *Nat. Commun.*, **5**, 5248.
37. Castello,A., Fischer,B., Frese,C.K., Horos,R., Alleaume,A.M., Foehr,S., Curk,T., Krijgsveld,J. and Hentze,M.W. (2016) Comprehensive identification of RNA-binding domains in human cells. *Mol. Cell*, **63**, 696–710.
38. Haberman,N., Huppertz,I., Attig,J., Konig,J., Wang,Z., Hauer,C., Hentze,M.W., Kulozik,A.E., Le Hir,H., Curk,T. *et al.* (2017) Insights into the design and interpretation of iCLIP experiments. *Genome Biol.*, **18**, 7.
39. Bradrick,S.S. and Gromeier,M. (2009) Identification of gemin5 as a novel 7-methylguanosine cap-binding protein. *PLoS One*, **4**, e7030.
40. Pineiro,D., Fernandez-Chamorro,J., Francisco-Velilla,R. and Martinez-Salas,E. (2015) Gemin5: A multitasking RNA-Binding protein involved in translation control. *Biomolecules*, **5**, 528–544.
41. Kim,M.S., Pinto,S.M., Getnet,D., Nirujogi,R.S., Manda,S.S., Chaerkady,R., Madugundu,A.K., Kelkar,D.S., Isserlin,R., Jain,S. *et al.* (2014) A draft map of the human proteome. *Nature*, **509**, 575–581.
42. Uhlen,M., Fagerberg,L., Hallstrom,B.M., Lindskog,C., Oksvold,P., Mardinoglu,A., Sivertsson,A., Kampf,C., Sjostedt,E., Asplund,A. *et al.* (2015) Proteomics. Tissue-based map of the human proteome. *Science*, **347**, 1260419.
43. Gates,J., Lam,G., Ortiz,J.A., Losson,R. and Thummel,C.S. (2004) rigor mortis encodes a novel nuclear receptor interacting protein required for ecdysone signaling during Drosophila larval development. *Development*, **131**, 25–36.